

# METHOD FOR SCREENING FOR DRUG CANDIDATES FOR MODULATING TRANSCRIPTION FACTOR ACTIVITY

Inventor: Jason Li

## 5 Field of the Invention

The present invention relates to methods for detecting activated transcription factors in a cell sample. More specifically, the invention relates to methods for detecting multiple different activated transcription factors in a cell sample at the  
10 same time and uses arising there from.

## Description of Related Art

All living organisms use nucleic acids (DNA and RNA) to encode the genes which make up the genome for that organism. Each gene encodes a protein that may  
15 be produced by the organism through expression of the gene.

It is important to note that the mere presence of a gene in a cell does not communicate the functionality of that gene to the cell. Rather, it is only when the gene is expressed and a protein is produced that the functionality of the gene encoding the protein is conveyed.

20 The systems that regulate gene expression respond to a wide variety of developmental and environmental stimuli, thus allowing each cell type to express a unique and characteristic subset of its genes, and to adjust the dosage of particular gene products as needed. The importance of dosage control is underscored by the fact that targeted disruption of key regulatory molecules in mice often results in  
25 drastic phenotypic abnormalities [Johnson, R. S., et al., Cell, 71:577-586 (1992)], just as inherited or acquired defects in the function of genetic regulatory mechanisms contribute broadly to human disease.

The importance of controlled gene expression in human disease and the information available to date relating to the mechanisms of gene regulation have  
30 fueled efforts aimed at discovering ways of overriding endogenous regulatory controls or of creating new signaling circuitry in cells [Belshaw, P. J., et al., Proc. Natl. Acad. Sci. USA, 93:4604-4607 (1996); Ho, S. H., et al., Nature (London),

382:822-826 (1996); Rivera, V. M., et al., Nat. Med., 2:1028-1032; Spencer, D. M., et al., Science, 262:1019-1024 (1993)].

Critical to this research are effective tools for monitoring gene expression. It is therefore of interest to be able to rapidly and accurately determine the relative expression of different genes in different cells, tissues and organisms, over time, and under various conditions, treatments and regimes. As will be described herein in greater detail, there are a great many applications that arise from being able to effectively monitor which genes are being expressed by a given cell at a given time.

Standard molecular biology techniques have been used to analyze the expression of genes in a cell by measuring DNA. These techniques include PCR, northern blot analysis, or other types of DNA probe analysis such as *in situ* hybridization. Each of these methods allows one to analyze the transcription of only known genes and/or small numbers of genes at a time. Nucl. Acids Res. 19, 7097-7104 (1991); Nucl. Acids Res. 18, 4833-4842 (1990); Nucl. Acids Res. 18, 2789-2792 (1989); European J. Neuroscience 2, 1063-1073 (1990); Analytical Biochem. 187, 364-373 (1990); Genet. Annal Techn. Appl. 7, 64-70 (1990); GATA 8(4), 129-133 (1991); Proc. Natl. Acad. Sci. USA 85, 1696-1700 (1988); Nucl. Acids Res. 19, 1954 (1991); Proc. Natl. Acad. Sci. USA 88, 1943-1947 (1991); Nucl. Acids Res. 19, 6123-6127 (1991); Proc. Natl. Acad. Sci. USA 85, 5738-5742 (1988); Nucl. Acids Res. 16, 10937 (1988).

Gene expression has also been monitored by measuring levels of mRNA. Since proteins are transcribed from mRNA, it is possible to detect transcription by measuring the amount of mRNA present. One common method, called "hybridization subtraction", allows one to look for changes in gene expression by detecting changes in mRNA expression. Nucl. Acids Res. 19, 7097-7104 (1991); Nucl. Acids Res. 18, 4833-4842 (1990); Nucl. Acids Res. 18, 2789-2792 (1989); European J. Neuroscience 2, 1063-1073 (1990); Analytical Biochem. 187, 364-373 (1990); Genet. Annal Techn. Appl. 7, 64-70 (1990); GATA 8(4), 129-133 (1991); Proc. Natl. Acad. Sci. USA 85, 1696-1700 (1988); Nucl. Acids Res. 19, 1954 (1991); Proc. Natl. Acad. Sci. USA 88, 1943-1947 (1991); Nucl. Acids Res. 19,

6123-6127 (1991); Proc. Natl. Acad. Sci. USA 85, 5738-5742 (1988); Nucl. Acids Res. 16, 10937 (1988).

Gene expression has also been monitored by measuring levels of gene product, (i.e., the expressed protein), in a cell, tissue, organ system, or even organism. Measurement of gene expression by measuring the protein gene product may be performed using antibodies known to bind to a particular protein to be detected. A difficulty arises in needing to generate antibodies to each protein to be detected. Measurement of gene expression via protein detection may also be performed using 2-dimensional gel electrophoresis, wherein proteins can be, in principle, identified and quantified as individual bands, and ultimately reduced to a discrete signal. In order to positively analyze each band, each band must be excised from the membrane and subjected to protein sequence analysis using Edman degradation. Unfortunately, it tends to be difficult to isolate a sufficient amount of protein to obtain a reliable sequence. In addition, many of the bands contain more than one discrete protein.

A further difficulty associated with quantifying gene expression by measuring an amount of protein gene product in a cell is that protein expression is an indirect measure of gene expression. It is impossible to know from a protein present in a cell when that protein was expressed by the cell. As a result, it is hard to determine whether protein expression changes over time due to cells being exposed to different stimuli.

Gene expression has also been monitored by measuring the amount of particular activated transcription factors present in a cell. Transcription in a cell is controlled by proteins, referred to herein as "activated transcription factors" which bind to DNA at sites outside the core promoter for the gene and activate transcription. Since activated transcription factors activate transcription, detection of their presence is useful for measuring gene expression. Transcriptional activators are found in prokaryotes, viruses, and eukaryotes, including fungi, plants, and animals, including mammals, providing a wide range of therapeutic targets.

The regulatory mechanisms controlling the transcription of protein-coding genes by RNA polymerase II have been extensively studied. RNA polymerase II

and its host of associated proteins are recruited to the core promoter through non-covalent contacts with sequence-specific DNA binding proteins [Tjian, R. and Maniatis, T., *Cell*, 77:5-8 (1994); Stringer, K. F., *Nature (London)*, 345:783-786 (1990)]. An especially prevalent and important subset of such proteins, known as transcription factors, typically bind DNA at sites outside the core promoter and activate transcription through space contacts with components of the transcriptional machinery, including chromatin remodeling proteins [Tjian, R. and Maniatis, T., *Cell*, 77:5-8 (1994); Stringer, K. F., *Nature (London)*, 345:783-786 (1990); Bannister, A. J. and Kouzarides, T., *Nature*, 384:641-643 (1996); Mizzen, C. A., et al., *Cell*, 87:1261-1270 (1996)]. The DNA-binding and activation functions of transcription factors generally reside on separate domains whose operation is portable to heterologous fusion proteins [Sadowski, I., et al., *Nature*, 335:563-564 (1988)]. Though it is believed that activation domains are physically associated with a DNA-binding domain to attain proper function, the linkage between the two need not be covalent [Belshaw, P. J., et al., *Proc. Natl. Acad. Sci. USA*, 93:4604-4607 (1996); Ho, S. H., et al., *Nature (London)*, 382:822-826 (1996)]. In many instances, the activation domain does not appear to contact the transcriptional machinery directly, but rather through the intermediacy of adapter proteins known as coactivators [Silverman, N., et al., *Proc. Natl. Acad. Sci. USA*, 91:11005-11008 ((1994); Arany, Z., et al., *Nature (London)*, 374:81-84 (1995)].

One of the difficulties associated with measuring gene expression by measuring transcription factors is that one must measure the subset of transcription factors which are "activated." Certain post-transcriptional modifications occur which render transcription factors "active" in the sense that they are capable of binding to DNA. It is thus necessary to distinguish between activated and non-activated transcription factors so that the "activated transcription factors" can be selectively measured.

Several different methods have been developed for detecting activated transcription factors. One method involves using antibodies selective for activated transcription factors over inactive forms of the transcription factor. This method is impractical for detecting multiple different activated transcription factors due to

difficulties associated with developing numerous different antibodies having the requisite bind specificities.

Another method for detecting activated transcription factors involves measuring DNA - transcription factor complexes through a gel shift assay.

5 [Ausebel, F.M. et al eds (1993) Current Protocols in Molecular Biology Vol.2  
Greene Publishing Associates, Inc. and John Wiley and Sons, Inc., New York].  
According to this method, a sample containing an activated transcription factor is  
contacted with a DNA probe that comprises a recognition sequence for the  
transcription factor. A complex between the activated transcription factor and the  
10 DNA probe is formed. The DNA-protein complex is detected by a gel-shift assay.  
Since individual gel shift assays must be performed for each activated transcription  
factor - DNA complex, this method is currently impractical for measuring multiple  
different activated transcription factors at the same time.

U.S. Patent Nos. 6,066,452 and 5,861,246 describe methods for determining  
15 DNA binding sites for DNA-binding proteins. The DNA binding sites may then be  
used as probes to isolate DNA-binding proteins. Similarly, PCT Publication No.  
WO 00/04196 describes methods for identifying cis acting nucleic acid elements as  
well as methods for isolating nucleic acid binding factors.

## 20 SUMMARY OF THE INVENTION

The present invention relates to methods and kits for isolating DNA probes  
that bind to activated transcription factors.

In one embodiment, a method is provided which comprises: contacting a  
25 biological sample with a library of double stranded DNA probes under conditions  
where DNA probe - transcription factor complexes are formed between the DNA  
probes and activated transcription factors present in the biological sample;  
separating DNA probe - transcription factor complexes from non-complexed DNA  
probes in the library using an agarose gel separation; excising a portion of the  
30 agarose gel comprising the separated DNA probe - transcription factor complexes;  
and isolating the DNA probes from the excised portion of the agarose gel.

In another embodiment, a method is provided which comprises: contacting a biological sample with a library of double stranded DNA probes under conditions where DNA probe - transcription factor complexes are formed between the DNA probes and activated transcription factors present in the biological sample;

5 separating DNA probe - transcription factor complexes from non-complexed DNA probes in the library using an agarose gel separation; excising a portion of the agarose gel comprising the separated DNA probe - transcription factor complexes; isolating the DNA probes from the excised portion of the agarose gel; and identifying which of the DNA probes in the library are isolated.

10 In another embodiment, a kit is provided which comprises: a library of double stranded DNA probes, each probe comprising a recognition sequence to which an activated transcription factor is capable of binding and forming a DNA probe - transcription factor complex, the DNA probes in the library capable of forming DNA probe - transcription factor complexes with multiple different  
15 activated transcription factors; and instructions for separating DNA probe - transcription factor complexes from non-complexed DNA probes in the library by agarose gel separation.

Kits are also provided for DNA probe libraries for detecting activated transcription factors.

20 In one embodiment, the kit comprises: first and second libraries of double stranded DNA probes, each probe in the first and second libraries comprising a recognition sequence to which an activated transcription factor is capable of binding and forming a DNA probe - transcription factor complex, the DNA probes in the library capable of forming DNA probe - transcription factor complexes with  
25 multiple different activated transcription factors; wherein the probes of the first library further comprise a first detectable marker and the probes of the second library further comprise a second detectable marker that is different than the first detectable marker.

Methods, arrays and kits are also provided for detecting activated  
30 transcription factors using a hybridization array.



In one embodiment, a method is provided which comprises: taking a library of double stranded transcription factor probes, the transcription factor probes each comprising a recognition sequence capable of binding to an activated transcription factor, the recognition sequence varying within the library for binding to different activated transcription factors; contacting a biological sample with the library of double stranded DNA probes under conditions where DNA probe - transcription factor complexes are formed between the DNA probes and activated transcription factors present in the biological sample; isolating the transcription factor probes from the transcription factor probe - transcription factor complexes formed; and identifying which transcription factor probes in the library formed complexes by taking an array of immobilized hybridization probes capable of hybridizing to at least one of the strands of the different double stranded transcription factor probes in the library and contacting the isolated transcription factor probes with the array under conditions suitable for hybridization of the strands of the different double stranded transcription factor probes to the hybridization probes in the array.

In another embodiment, a hybridization array is provided for use in identifying which of a plurality of different activated transcription factors are present in a biological sample by immobilizing transcription factor probes that form transcription factor probe - transcription factor complexes with different activated transcription factors, the array comprising: a substrate; and a plurality of hybridization probes immobilized on a surface of the substrate such that different hybridization probes are positioned in different defined regions on the surface, the different hybridization probes comprising a different transcription factor probe binding region capable of immobilizing a different transcription factor probe to the array, the transcription factor probe binding region comprising at least two copies of a complement to a portion of a recognition sequence comprised on the transcription factor probe. The hybridization array may optionally further comprise an internal standard. For example, the array may further comprise biotinylated DNA which is employed as an internal standard.

In another embodiment, a kit is provided for use in identifying which of a plurality of different activated transcription factors are present in a biological sample

by isolating and immobilizing transcription factor probes that form transcription factor probe - transcription factor complexes with different activated transcription factors, the kit comprising: a hybridization array comprising a substrate, and a plurality of hybridization probes immobilized on a surface of the substrate such that different hybridization probes are positioned in different defined regions on the surface, the different hybridization probes comprising a different transcription factor probe binding region capable of immobilizing a different transcription factor probe to the array, the transcription factor probe binding region comprising at least two copies of a complement to a portion of a recognition sequence comprised on the transcription factor probe; and instructions for separating DNA probe - transcription factor complexes from non-complexed DNA probes in the library by agarose gel separation.

Methods for characterizing cell types based on which activated transcription factors are present in a sample are also provided.

In one embodiment, a method is provided which comprises: taking a library of double stranded transcription factor probes, the transcription factor probes each comprising a recognition sequence capable of binding to an activated transcription factor, the recognition sequence varying within the library for binding to different activated transcription factors native to different cell types; contacting a biological sample with the library of double stranded DNA probes under conditions where DNA probe - transcription factor complexes are formed between the DNA probes and activated transcription factors present in the biological sample; isolating the transcription factor probes from the transcription factor probe - transcription factor complexes formed; identifying which transcription factor probes in the library formed complexes by taking an array of immobilized hybridization probes capable of hybridizing to at least one of the strands of the different double stranded transcription factor probes in the library and contacting the isolated transcription factor probes with the array under conditions suitable for hybridization of the strands of the different double stranded transcription factor probes to the hybridization probes in the array; and identifying a cell type of the biological sample based on which transcription factor probes are identified.



Methods for identifying a disease state based on which activated transcription factors are present in a biological sample are also provided.

In one embodiment, the method comprises taking a library of double stranded transcription factor probes, the transcription factor probes each comprising  
5 a recognition sequence capable of binding to an activated transcription factor, the recognition sequence varying within the library for binding to different activated transcription factors native to different cell types; identifying which activated transcription factors are present in a nuclear extract of a test sample of cells by:  
contacting the nuclear extract of the test sample with the library of double stranded  
10 DNA probes under conditions where DNA probe - transcription factor complexes are formed between the DNA probes and activated transcription factors present in the test sample, isolating the transcription factor probes from the transcription factor probe - transcription factor complexes formed, and identifying which transcription factor probes in the library formed complexes by taking an array of immobilized  
15 hybridization probes capable of hybridizing to at least one of the strands of the different double stranded transcription factor probes in the library and contacting the isolated transcription factor probes with the array under conditions suitable for hybridization of the strands of the different double stranded transcription factor probes to the hybridization probes in the array; identifying which activated  
20 transcription factors are present in a nuclear extract of a control sample of cells by:  
contacting the nuclear extract of the control sample with the library of double stranded DNA probes under conditions where DNA probe - transcription factor complexes are formed between the DNA probes and activated transcription factors present in the control sample, isolating the transcription factor probes from the  
25 transcription factor probe - transcription factor complexes formed, and identifying which transcription factor probes in the library formed complexes by taking an array of immobilized hybridization probes capable of hybridizing to at least one of the strands of the different double stranded transcription factor probes in the library and contacting the isolated transcription factor probes with the array under conditions  
30 suitable for hybridization of the strands of the different double stranded transcription

factor probes to the hybridization probes in the array; and comparing which activated transcription factors are present in the test sample and the control sample.

Methods for screening drug candidates for modulating an activated transcription factor's activity are also provided.

5 In one embodiment, the method comprises: forming a plurality of test samples by contacting samples of cells with different agents; and for each test sample, identifying which of a plurality of different activated transcription factors are present by: taking a library of double stranded transcription factor probes, the transcription factor probes each comprising a recognition sequence capable of  
10 binding to an activated transcription factor, the recognition sequence varying within the library for binding to different activated transcription factors, contacting the different test sample with the library of double stranded DNA probes under conditions where DNA probe - transcription factor complexes are formed between the DNA probes and activated transcription factors present in the test samples,  
15 isolating the transcription factor probes from the transcription factor probe - transcription factor complexes formed, and identifying which transcription factor probes in the library formed complexes by taking an array of immobilized hybridization probes capable of hybridizing to at least one of the strands of the different double stranded transcription factor probes in the library and contacting the  
20 isolated transcription factor probes with the array under conditions suitable for hybridization of the strands of the different double stranded transcription factor probes to the hybridization probes in the array; and comparing the activated transcription factors present in the different test samples.

Methods for determining sequence binding requirements for an activated  
25 transcription factor are also provided.

In one embodiment, the method comprises: contacting a sample comprising an activated transcription factor with a library of double stranded DNA probes under conditions where DNA probe - transcription factor complexes are formed between the DNA probes and the activated transcription factor; separating DNA probe -  
30 transcription factor complexes from non-complexed DNA probes in the library; isolating the DNA probes from the excised portion of the agarose gel; and

determining a consensus sequence for the DNA probes isolated in order to assess the binding requirements for the transcription factor.

In another embodiment, the method comprises contacting a sample comprising an activated transcription factor with a library of double stranded DNA probes under conditions where DNA probe - transcription factor complexes are formed between the DNA probes and the activated transcription factor;

separating DNA probe - transcription factor complexes from non-complexed DNA probes in the library using an agarose gel separation; excising a portion of the agarose gel comprising the separated DNA probe - transcription factor complexes; isolating the DNA probes from the excised portion of the agarose gel; and determining a consensus sequence for the DNA probes isolated in order to assess the binding requirements for the transcription factor.

In yet another embodiment, the method comprises: contacting a sample comprising an activated transcription factor with a library of double stranded DNA probes under conditions where DNA probe - transcription factor complexes are formed between the DNA probes and the activated transcription factor;

separating DNA probe - transcription factor complexes from non-complexed DNA probes in the library; isolating the DNA probes from the excised portion of the agarose gel; and quantifying the amount of each of the isolated DNA probes.

In yet another embodiment, the method comprises: contacting a sample comprising an activated transcription factor with a library of double stranded DNA probes under conditions where DNA probe - transcription factor complexes are formed between the DNA probes and the activated transcription factor; separating DNA probe - transcription factor complexes from non-complexed DNA probes in the library using an agarose gel separation; excising a portion of the agarose gel comprising the separated DNA probe - transcription factor complexes; isolating the DNA probes from the excised portion of the agarose gel; and quantifying the amount of each of the isolated DNA probes.

Methods are also provided for quantifying expression and activation of multiple different activated transcription factors. According to one embodiment, the method comprises: contacting a biological sample with a library of double stranded

DNA probes for detecting active forms of multiple different transcription factors under conditions where DNA probe - transcription factor complexes are formed between the DNA probes and activated transcription factors present in the biological sample; isolating DNA probes from the DNA probe - transcription factor complexes; identifying which of the multiple different transcription factors are present in an activated form in the biological sample based on which DNA probes are isolated; and quantifying expression of the multiple different transcription factors from cDNA for the biological sample.

## 10 BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 provides a flow diagram for a method for identifying which of a plurality of transcription factors are activated in a given sample of cells.

15 Figure 2 illustrates an array of hybridization probes attached to a solid support where different hybridization probes are attached to discrete, different regions of the array. A transcription factor expression signature is shown based on the distribution of where transcription factor probes hybridize and their intensity.

Figure 3 illustrates an array of hybridization probes attached to a solid support where probes from a first sample with a first color dye and probes from a second sample with a second color dye are both contacted with the array.

Figure 4 illustrates a process whereby the minimum DNA sequence binding requirements for a given transcription factor can be rapidly determined.

Figure 5 illustrates a variation of the method described in regard to Figure 4 where an optimal sequence for binding is identified.

25 Figure 6 provides the sequences for the probes used to form the transcription factor probe library used in the experiments described in Sections 12-19 herein.

Figure 7 depicts the layout of the array of hybridization probes employed in the experiments described in Sections 12-19 herein.

30 Figure 8 is an image of the array described in regard to Figure 7 when the transcription factor probe library described in Figure 6 is contacted with the array.

Figure 9A is an image of an array described in regard to Figure 7 that is contacted with transcription factor hybridization probes isolated from transcription factor probe - transcription factor complexes formed when Brn3 transcription factor hybridization probes are contacted with a nuclear extract of HeLa cells.

5        Figure 9B is an image of an array described in regard to Figure 7 that is contacted with transcription factor hybridization probes isolated from transcription factor probe - transcription factor complexes formed when c-Myb transcription factor hybridization probes are contacted with a nuclear extract of HeLa cells.

10       Figure 9C is an image of an array described in regard to Figure 7 that is contacted with transcription factor hybridization probes isolated from transcription factor probe - transcription factor complexes formed when Smad3/4 transcription factor hybridization probes are contacted with a nuclear extract of HeLa cells.

15       Figure 9D is an image of an array described in regard to Figure 7 that is contacted with transcription factor hybridization probes isolated from transcription factor probe - transcription factor complexes formed when Brn3, c-Myb, and Smad3/4 transcription factor hybridization probes are contacted with a nuclear extract of HeLa cells.

20       Figure 10A is an image of an array described in regard to Figure 7 that is contacted with transcription factor hybridization probes isolated from transcription factor probe - transcription factor complexes formed when the entire library of transcription factor probes described in regard to Figure 6 are contacted with a control sample that does not contain any transcription factors.

25       Figure 10B is an image of an array described in regard to Figure 7 that is contacted with transcription factor hybridization probes isolated from transcription factor probe - transcription factor complexes formed when the entire library of transcription factor probes described in regard to Figure 6 are contacted with a nuclear extract of HeLa cells.

30       Figure 11A is an image of an array described in regard to Figure 7 that is contacted with transcription factor hybridization probes isolated from transcription factor probe - transcription factor complexes formed when the entire library of

transcription factor probes described in regard to Figure 6 are contacted with a nuclear extract of HeLa cells.

Figure 11B is an image of an array described in regard to Figure 7 that is contacted with transcription factor hybridization probes isolated from transcription factor probe - transcription factor complexes formed when the entire library of transcription factor probes described in regard to Figure 6 are contacted with a nuclear extract of PMA-treated HeLa cells.

Figure 12A provides a table of the signal intensities of regions of the array shown in Figure 11A.

Figure 12B provides a table of the signal intensities of regions of the array shown in Figure 11B.

Figure 12C provides a table with the ratios between the intensities of the regions of the arrays shown in Figures 12A and 12B.

Figure 13 provides an image of a gel shift analysis of Est and NF-E1 performed on HeLa and PMA-treated HeLa cells.

Figure 14A is an image of an array described in regard to Figure 7 that is contacted with transcription factor hybridization probes isolated from transcription factor probe - transcription factor complexes formed when the entire library of transcription factor probes described in regard to Figure 6 are contacted with a nuclear extract of A431 cells.

Figure 14B is an image of an array described in regard to Figure 7 that is contacted with transcription factor hybridization probes isolated from transcription factor probe - transcription factor complexes formed when the entire library of transcription factor probes described in regard to Figure 6 are contacted with a nuclear extract of PMA-treated A431 cells.

Figure 15 provides an image of a gel shift analysis of Ets, NF-E1, and NF-kB performed on A431 and PMA-treated A431 cells.

Figure 16A is an image of an array described in regard to Figure 7 that is contacted with transcription factor hybridization probes isolated from transcription factor probe - transcription factor complexes formed when the entire library of



transcription factor probes described in regard to Figure 6 are contacted with a nuclear extract of Jurkat cells.

Figure 16B is an image of an array described in regard to Figure 7 that is contacted with transcription factor hybridization probes isolated from transcription factor probe - transcription factor complexes formed when the entire library of transcription factor probes described in regard to Figure 6 are contacted with a nuclear extract of PMA-treated Jurkat cells.

Figure 17A is an image of an array described in regard to Figure 7 that is contacted with transcription factor hybridization probes isolated from transcription factor probe - transcription factor complexes formed when the entire library of transcription factor probes described in regard to Figure 6 are contacted with a nuclear extract of HeLa cells.

Figure 17B is an image of an array described in regard to Figure 7 that is contacted with transcription factor hybridization probes isolated from transcription factor probe - transcription factor complexes formed when the entire library of transcription factor probes described in regard to Figure 6 are contacted with a nuclear extract of A431 cells.

Figure 17C is an image of an array described in regard to Figure 7 that is contacted with transcription factor hybridization probes isolated from transcription factor probe - transcription factor complexes formed when the entire library of transcription factor probes described in regard to Figure 6 are contacted with a nuclear extract of Jurkat cells.

Figure 17D is an image of an array described in regard to Figure 7 that is contacted with transcription factor hybridization probes isolated from transcription factor probe - transcription factor complexes formed when the entire library of transcription factor probes described in regard to Figure 6 are contacted with a nuclear extract of K-562 cells.

Figure 17E is an image of an array described in regard to Figure 7 that is contacted with transcription factor hybridization probes isolated from transcription factor probe - transcription factor complexes formed when the entire library of

transcription factor probes described in regard to Figure 6 are contacted with a nuclear extract of Y79 cells.

Figure 18A is an image of a polyacrylamide gel from a gel shift analysis for multiple different transcription factors.

5        Figure 18B is an image of an agarose gel from a gel shift analysis for multiple different transcription factors.

#### DETAILED DESCRIPTION OF THE INVENTION

10        The present invention relates to rapid and efficient methods for identifying multiple different activated transcription factors in a biological sample at the same time. The methods of the present invention also provide for the quantification of the multiple different activated transcription factors. As will be described herein in greater detail, there are a great many applications that arise from being able to  
15        effectively monitor which genes are being expressed at a given time. With the assistance of the methods of the present invention, it is thus possible to rapidly and effectively monitor the levels of expression of multiple different genes at the same time.

20        The present invention also relates to various compositions, kits, and devices for use in conjunction with the various methods of the present invention.

Figure 1 provides a general overview of how the present invention detects multiple different activated transcription factors and thus allows the expression of multiple different genes to be simultaneously monitored.

25        As illustrated, a biological sample is contacted with a library of probes. The biological sample is typically derived from a sample of cells and is preferably a nuclear extract of the cell sample. The sample contains an unknown mixture of activated transcription factors. Which activated transcription factors are present in the sample serves to indicate which genes are currently being expressed.

30        Information about the DNA binding specificity of transcription factors which one wishes to identify is used to design a library of transcription factor probes. In this regard, it is noted that the present invention relates to the detection, monitoring and optionally quantification of known transcription factors using a library of

transcription factor probes whose sequences are known. The probes in the library are preferably double stranded. One of the strands is preferably biotin labeled at the 5' end to facilitate detection.

Each transcription factor probe in the library comprises a DNA sequence that is capable of binding to an activated transcription factor, referred to herein as the probe's recognition sequence. At least the recognition sequence varies within the library of probes such that the probes are capable of binding to a plurality of different activated transcription factors. Due to the high level binding specificity of the transcription factors, each transcription factor probe has binding specificity for a single transcription factor or family of transcription factors.

Because the present invention is used to identify known transcription factors, it is practical to use longer recognition sequences in the probes in the library as compared to what would be practical if a random library were used. As a result, at least 1%, 2%, 3%, 5%, 10%, 20%, 30%, 50% or more of the probes in the library may have recognition sequences greater than 35, 40, 45 or more base pairs in length.

By using longer recognition sequences, the probes have greater binding specificity. In addition, the probes have greater binding efficiency to the transcription factors which improves the yield of probe – transcription factor complexes isolated. As a result, the method of the present invention provides a high level of sensitivity for isolating probe – transcription factor complexes, as described further herein, in combination with a high signal to noise ratio.

As a result of contacting the sample with the library of transcription factor probes, complexes are formed between activated transcription factors present in the sample and transcription factor probes in the library which have sequences that match the sequence specificity of the DNA-binding domains of the activated transcription factors. By isolating the transcription factor probe – activated transcription factor complexes, those probes from the library which bind to transcription factors in the sample are isolated.

The isolated transcription factor probes are then identified. Each probe is specific for a different transcription factor. Since only those probes from the library which form a complex with an activated transcription factor will be isolated,

identification of which probes are isolated serves to identify which activated transcription factors are present. Since the presence of an activated transcription factor evidences gene expression, the above described method can be used to determine which genes were being expressed at the time the sample was taken based on which activated transcription factors are present.

The design, operation and applications for the present invention will now be described in greater detail.

# **1. Libraries of Transcription Factor Probes**

10

Libraries of transcription factor probes are provided that may be used to detect activated transcription factors according to the present invention. A given library comprises a plurality of double stranded DNA probes where the DNA sequences of the probes vary within the library. The DNA sequences employed in the probes of the libraries preferably have a length between about 10 and 100 base pairs, preferably between about 10 and 75 base pairs, more preferably between about 15 and 50. Probes of longer lengths may also be used.

Each probe in the library comprises a recognition sequence which is capable of forming a probe – transcription factor complex with an activated transcription factor. Due to the high level of DNA binding specificity of transcription factors, each transcription factor will typically bind to a different DNA sequence. In some instances, a related family of transcription factors may bind to the same DNA sequence.

By designing the library of probes such that any given probe in the library includes a DNA recognition sequence which a particular activated transcription factor (or a related family of activated transcription factors) will bind to, and does not also include a DNA recognition sequence which other activated transcription factors will bind to, a given probe may be used to identify a single activated transcription factor (or a related family of activated transcription factors).

It is noted that in certain situations, individual probes may bind to more than one activated transcription factor and may nonetheless be used in the library. For

example, certain probes may bind to a related family of activated transcription factors. Less than 1:1 binding specificity between probes and activated transcription factors can be readily resolved during analysis of the isolated probes.

5 The DNA recognition sequences used in the probes in the libraries preferably have a length between about 10 and 100 base pairs, more preferably between about 10 and 75 base pairs, more preferably between about 15 and 50 base pairs, more preferably between about 20 and 40 base pairs, and most preferably a length between about 25 and 35 base pairs.

10 The optimal length for the recognition sequence and the overall probe may vary somewhat depending on the particular transcription factor. Hence, one may wish to evaluate the optimal length for the recognition sequence and the probe for a given transcription factor using a traditional gel shift assay.

15 Because the present invention is used to identify known transcription factors, it is practical to use longer recognition sequences in the probes in the library as compared to what would be practical if a random library were used. For example, at least 1%, 2%, 3%, 5%, 10%, 20%, 30%, 50% or more of the probes in the library may have recognition sequences greater than 35, 40, 45 or more base pairs in length. By using longer recognition sequences, the probes have greater binding specificity and greater binding efficiency to the transcription factors. As a result, the method of  
20 the present invention provides a high level of sensitivity for isolating probe - transcription factor complexes, as described further herein, in combination with a high signal to noise ratio.

Selection of which DNA recognition sequences to use in a library may be based on the different transcription factors that one wishes to detect in a sample.  
25 This, in turn, may depend on the type of organism, cell, or disease state one wishes to identify and/or monitor the gene expression of. It may also depend on the different functionality that one wishes to identify or monitor.

A significant feature of the present invention is the ability to detect multiple different transcription factors at the same time. This ability arises from the number  
30 of different DNA recognition sequences used in a library, the number of different DNA recognition sequences relating directly to the number of different transcription

factors that the library can be used to detect. A given library of transcription factor probes preferably has at least 2, 3, 5, 10, 20, 50, 100, 250, or more different DNA recognition sequences. The upper limit on the number of different DNA recognition sequences that may be incorporated into a library is limited only by the number of  
5 known DNA recognition sequences.

A given library of transcription factor probes may be used to detect gene expression in a single type of cell or organism or may be used to detect gene expression in multiple different types of cells or organisms. When the library is designed to detect gene expression in multiple different types of cells or organisms,  
10 the library has DNA recognition sequences for multiple different types of cells or organisms. For example, the library may include DNA recognition sequences for 2, 3, 4, 5 or more different types of cells or organisms. In one embodiment, the library includes DNA recognition sequences for 10, 20, 30, 50, or more different types of cells or organisms.

15 If the sample comprises cells that may be from one or more different organisms, the DNA recognition sequences used in the library may be for all or some of the different transcription factors expressed by the one or more different organisms. For example, if the library is to be used to classify an unknown type of bacterium, the library may include DNA recognition sequences for multiple different  
20 types of bacteria, thereby allowing the library to be used to classify the bacterium.

If the sample comprises cells from a particular organism, the DNA recognition sequences used in the library may be for all or some of the different transcription factors expressed by organism. If the library is to be used to classify an unknown type of cells (i.e., determine whether a growth is malignant), the library  
25 may include DNA recognition sequences for multiple different types of cells including the different types of malignant, benign, and normal cell types present in the organism.

If the sample comprises cells of a single cell type, the DNA recognition sequences used in the library may be for all or some of the different transcription  
30 factors expressed by that cell type. For example, if one wishes to monitor the expression of only a particular group of genes, such as the genes associated with a



particular pathway, the library may include DNA recognition sequences for the transcription factors associated with that group of genes.

As one can see, a myriad of different libraries of probes can be assembled depending on the nature of the sample and the nature of the analysis to be performed. It is noted that different libraries may also be formed when a particularly large number of transcription factors are to be detected or when different binding conditions are needed for different groups of probes.

The probes in the library may optionally further comprise a detectable marker which allows the probe to be detected once isolated from the transcription factor probe – transcription factor complex. Since a wide variety of detection techniques may be used to identify the isolated probes, a similarly wide range of detectable markers may be used in conjunction with those different detection techniques.

The detectable marker may be any marker which can be used to determine the presence or absence of the DNA probes. In a preferred embodiment, the detectable marker is biotin and is preferably attached to one of the 5' ends of probes. Biotin probes have been found to provide a desirable high level of sensitivity.

The detectable marker may also be a dye which can be seen under natural light or with the assistance of an excitation light source to cause fluorescence. In one embodiment, the detectable marker is a fluorescent dye. Examples of fluorescent dyes that may be used include, but are not limited to fluorescein and its derivatives, rhodamine and its derivatives, dansyl, umbeliferone, acridinium, and chemiluminescent molecules such as luciferin and 2,3-dihydrophthalazinediones. The fluorescent dye may also be an energy transfer fluorescent dye.

The detectable marker may also be a molecule which binds to an analytically detectable counterpart. For example, the detectable marker may be covalently attached to or incorporated into the substrate, for example, as taught by Ward, European Patent Application No. 63,879. In such instances, the substrate is detected by adding the analytically detectable counterpart which specifically binds to the substrate, thereby enabling detection of the substrate. Examples of such detectable markers and their analytically detectable counterparts include biotin and either

fluorescent or chemiluminescent avidin. Antibodies that bind to an analytically detectable antigen may also be used as the detectable marker. The detectable marker may also be a molecule which, when subjected to chemical or enzymatic modification, becomes detectable such as those disclosed in Leary, et al., Proc. Natl. Acad. Sci. (U.S.A.), 80:4045-4049 (1983).

In certain instances, it may be desirable to employ multiple different detectable markers. For example, if one wishes to classify an unknown type of cell or organism, the library may include DNA probes where different detectable markers are attached to the probes for the different types of cells. Hence, probes for transcription factors expressed by malignant cells may include a first fluorescent dye whereas probes expressed by benign cells may include a second fluorescent dye. In another example, probes for transcription factors expressed by a first type of lung cancer may include a first fluorescent dye whereas probes expressed by of a second, different type of lung cancer may include a second fluorescent dye. This allows one to rapidly visually identify the type of cell based on which detectable markers are present.

When one wishes to compare the gene expression of different groups of cells, multiple libraries may be prepared where each library contains a different detectable marker. For example, a first library labeled with a first detectable marker may be used with a first sample of cells and a second library labeled with a second detectable marker may be used with a second sample of cells. This way, the isolated probes may be analyzed together. As described in Section 6A, when an array format is used for detecting the isolated probes, the use of different detectable markers is particularly advantageous. For example, as shown in Figure 3, transcription factor probes isolated from complexes formed from a first sample (e.g., a control sample) may have a green dye, and transcription factor probes isolated from complexes formed from a second sample (e.g., a test sample) may have a red dye. Regions in the array which are green represent genes which only the cells in the control sample are expressing, regions in the array which are red represent genes which only cells in

the test sample are expressing, and regions in the array which are both green and red represent genes which cells in both the control and test samples are expressing.

## 2. Preparation of Sample

5

Nuclear extracts can be prepared from a sample of cells using the method described by Dignam, J.D., Lebovitz, R.M., and Roeder, R.G. (1983) Accurate transcription initiation by RNA polymerase II in a soluble extract from isolated mammalian nuclei. *Nucleic Acids Research* 11:1475–1489. Alternatively, a commercially available kit, such as Sigma's Nu-CLEAR Extraction Kit (cat. #N-TRACT) can be used.

## 3. Contacting Cell Sample With Library of Probes

15

Once a nuclear extract of a sample of cells is prepared, the nuclear extract is contacted with a library of probes and incubated at 15°C for 30 min.

## 4. Isolation of Probe – Transcription Factor Complexes

20

After contacting the sample with the library of transcription factor probes, any probe – activated transcription factor complexes formed are isolated. Any isolation method which can effectively isolate the complexes may be used. Isolation is preferably performed by a form of size separation, more preferably by electrophoresis, more preferably gel electrophoresis and most preferably gel electrophoresis using an agarose gel.

25

One of the problems overcome by the present invention is the ability to isolate complexes of DNA probes bound to transcription factors from other probes in the library. Several methods for performing this type of isolation were attempted. However, these methods failed to provide a sufficient yield of probe – transcription factor complexes from the sample.

30

For example, Applicant attempted to isolate probe – transcription factor complexes by performing an ammonia precipitation. Applicant also attempted to

isolate probe – transcription factor complexes by passing the sample through a nitrocellulose filter, the filter serving to immobilize proteins while allowing DNA that is not bound to protein to pass through the filter. Unfortunately, neither of these approaches provided a satisfactory yield of complexes for further characterization of the isolated probes.

Applicant also attempted to isolate probe – transcription factor complexes by using acrylamide gel electrophoresis. Unfortunately, this approach also did not provide a satisfactory yield of complexes. Without being bound by theory, it is believed that this method was hampered by the DNA probes being retained by the acrylamide gel.

Applicant successfully isolated probe – transcription factor complexes from the sample using agarose gel electrophoresis. Interestingly, despite the fact that agarose gel electrophoresis does not provide the same quality separation as other forms of gel electrophoresis (e.g., acrylamide gel electrophoresis), the resolution provided using agarose is more than sufficient to effectively separate the probe – transcription factor complexes. Meanwhile, agarose proved to have satisfactorily low retention of the DNA probes in the complex, thereby allowing the probes to be further characterized.

The separated probe – transcription factor complexes were isolated by removing the band from the gel containing the probe – transcription factor complexes.

It is noted that simultaneous isolation of multiple different probe – transcription factor complexes using a library of DNA probes is a significant departure from traditional gel-shift assays involving the use of a single probe to cause the gel shift of a single transcription factor. [Ausebel, F.M. et al eds (1993) Current Protocols in Molecular Biology Vol.2 Greene Publishing Associates, Inc. and John Wiley and Sons, Inc., New York ]. In such prior gel-shift assays, since only a single transcription factor was being detected, no efforts were made to isolate the probe used to cause the gel shift. By contrast, in the present invention, since multiple probes are used in combination it is necessary to isolate the complexes in

order to isolate and characterize which of the plurality of probes in the library formed a complex with a transcription factor and are present in the band.

It is also noted that simultaneous isolation of multiple different probe – transcription factor complexes by isolating and characterizing the DNA probes is also a significant departure from U.S. Patent Nos. 6,066,452 and 5,861,246 and PCT Publication No. WO 00/04196 which describe isolation of the nucleic acid binding factors, not the probes.

The following is a detailed description of how one may isolate probe – transcription factor complexes using an agarose gel. It is noted that other separation and isolation methods may be employed without departing from the present invention.

According to one embodiment, a 2% agarose gel in 0.5 X TBE is prepared. Preferably, 8 mm-wide combs are used. Each sample is mixed with 2 µl of a gel loading buffer. Table 1 provides an embodiment of a gel loading buffer that may be used.

**TABLE 1**

**Gel shift loading buffer:**

0.25 X TBE buffer: 60%  
Glycerol: 40%  
Bromophenol blue: 0.2% (w/v)

Add 2 ul of gel shift loading buffer and load into a 0.8 cm width lane of 2% agarose gel in 0.5% TBE. The resulting agarose gel is then run in 0.5% TBE at 120V for 16 min. The gel area containing that contains the protein/DNA complex, which will be above the blue dye and below the gel lane, is excised and transferred to a 1.5 mL tube.

## 5. **Separation of Transcription Factor Probes From Transcription Factor – Probe Complexes**

Once the transcription factor probe – transcription factor complexes are separated from other proteins and DNA in the sample, for example, as described in Section 4, the transcription factor probes are separated from the excised portion of the gel.

The following is a detailed description of how one may isolate probe – transcription factor complexes from an excised portion of an agarose gel. It is noted that other isolation techniques may be employed without departing from the present invention.

It is noted that the following steps describing the isolation of the DNA probes is specifically designed for CLONTECH's NucleoTrap Kit. If another commercially available gel extraction kit is employed, the steps should be modified in accordance with that manufacturer's instructions.

1.0 mL of NT1 solution from CLONTECH's NucleoTrap Kit is added to the excised portion of the gel. The mixture is then incubated at 50°C until the gel is totally dissolved. The tube is preferably periodically gently inverted in order to mix the contents until the gel is dissolved.

6 µl of beads from a commercially available gel extraction kit is then preferably added and the resulting mixture is incubated at room temperature for 10 min. The tube is preferably periodically gently inverted every 2–3 minutes.

The tube is then microfuged at 10,000 rpm for 30 sec. The resulting supernatant is carefully removed and the pellet resuspended in 150 µl of NT2 solution from CLONTECH's NucleoTrap Kit.

The pellet is then microfuged at 10,000 rpm for 30 sec. The resulting supernatant is then carefully removed and the pellet resuspended in 150 µl of NT3 solution from CLONTECH's NucleoTrap Kit. The pellet is then microfuged at 10,000 rpm for 30 sec. The supernatant is again removed and the pellet is allowed to air-dry for 10 min.



50  $\mu$ l of dH<sub>2</sub>O is then added to resuspend the pellet. The resulting mixture is incubated at room temperature for 5 min. The mixture is then gently shaken and incubated for another 5 min.

The resulting mixture is then microfuged at 10,000 rpm for 1 min. The supernatant is transferred to a fresh 1.5 ml tube. The isolated supernatant contains the DNA probes that are bound to proteins. The isolated supernatant is preferably stored on ice until proceeding to a characterization of the DNA probes.

## 6. Identifying Isolated Transcription Factor Probes

10

A variety of different methods may be used to identify which of the transcription factor probes from the library are present in the isolated probe – transcription factor complexes. These methods preferably also allow for the amount of transcription factor probes isolated to also be quantified. By identifying which transcription factor probes form complexes, one is able to determine which transcription factors are present in an activated form in the sample, the presence of an activated transcription factor evidencing expression of the gene associated with the transcription factor. By quantifying the amount of each transcription factor probe that forms a complex, one is able to determine the amount of each transcription factor present and hence the level of expression of the gene associated with that transcription factor.

One method that may be used to identify which of the transcription factor probes from the library are present in the isolated probe – transcription factor complexes is mass spectroscopy. According to this method, the length and composition of each probe can be determined. Therefore, the analyzed results show whether a specific probe is existing in the complexes and the interactions between transcription factors and binding probes can be determined.

Another method that may be used to identify which of the transcription factor probes from the library are present in the isolated probe – transcription factor complexes is based on size separation. According to this method, one varies the length of the probes in the transcription factor probe library so that it is possible to

resolve the different sized probes based on a size-based separation. For example, electrophoresis may be performed in order to separate the probes based on size. Such size-based DNA separations are traditionally done with high level specificity for DNA sequencing. By identifying which transcription factor probes are present  
5 based on the size-based separation, one can determine which activated transcription factors are present and can also quantify the amount of each activated transcription factor.

Yet another method for identifying which of the probes from the transcription factor probe library formed complexes involves hybridization of the  
10 transcription factor probes with a hybridization probe comprising a complement to the transcription factor probes recognition sequence. According to this method, detection of a particular transcription factor probe is accomplished by detecting the formation of a duplex between the transcription factor probe and a hybridization probe comprising a complement to the transcription factor probe's recognition  
15 sequence.

A wide variety of assays have been developed for performing hybridization assays and detecting the formation of duplexes that may be used in the present invention. For example, hybridization probes with a fluorescent dye and a quencher where the fluorescent dye is quenched when the probe is not hybridized to a target  
20 and is not quenched when hybridized to a target oligonucleotide may be used. Such fluorescer-quencher probes are described in, for example, U.S. Patent No. 6,070,787 and S. Tyagi et al., "Molecular Beacons: Probes that Fluoresce upon Hybridization", Dept. of Molecular Genetics, Public Health Research Institute, New York, N. Y., Aug. 25, 1995, each of which are incorporated herein by reference. By attaching  
25 different fluorescent dyes to different hybridization probes, it is possible to determine which transcription factor probes from the library formed complexes based on which fluorescent dyes are present (e.g, fluorescent dye and quencher on hybridization probe or fluorescent dye on hybridization probe and quencher on transcription factor probe). Applicant notes that one may also attach different  
30 fluorescent dyes to different transcription factor probes and use a change in fluorescence due to hybridization to a hybridization probe to determine which

transcription factor probes are present (e.g, fluorescent dye and quencher on transcription factor probe or fluorescent dye on transcription factor probe and quencher on hybridization probe).

5 A difficulty, however, arises when using multiple different fluorescer to detect multiple different transcription factor probes. Namely, there is a limited number of different fluorescers that may be spectrally resolved. As a result, a limited number of different transcription factors can be detected at the same time, for example only as many as five to ten.

#### 10                   A.       Hybridization Arrays For Detecting Isolated Transcription Factor Probes

15 A preferred assay for detecting the formation of duplexes between transcription factor probes and hybridization probes comprising their complements involves the use of an array of hybridization probes immobilized on a solid support. The hybridization probes comprise sequences that are complementary to at least a portion of the recognition sequences of the transcription factor probes and thus are able to hybridize to the different probes in a transcription factor probe library.

20 In order to improve enhance the sensitivity of the hybridization array, the immobilized probes preferably provide at least 2, 3, 4 or more copies of at least a portion of the recognition sequence incorporated into the transcription factor probes.

25 According to the present invention, the hybridization probes immobilized on the array preferably are at least 25 nucleotides in length, more preferably at least 30, 40 or 50 nucleotides in length. The immobilized hybridization probes may be 50, 75, 100 nucleotides or longer in length. In one preferred embodiment the immobilized probes are between 50 and 100 nucleotides in length.

30 By immobilizing hybridization probes on a solid support which comprise one or more copies of a complement to at least a portion of the recognition sequences of the transcription factor probes, the hybridization probes serve as immobilizing agents for the transcription factor probes, each different hybridization probe being designed to selectively immobilize a different transcription factor probe.

Figure 2 illustrates an array of hybridization probes attached to a solid support where different hybridization probes are attached to discrete, different regions of the array. Each different region of the array comprises one or more copies of a same hybridization probe which incorporates a sequence that is complementary to a recognition sequence of a transcription factor probe. As a result, the hybridization probes in a given region of the array can selectively hybridize to and immobilize a different transcription factor probe based on the transcription factor probe's recognition sequence.

By detecting which regions the isolated transcription factor probes hybridize to on the array, one can determine which activated transcription factors are present in the sample and can also quantify the amount of each activated transcription factor.

These arrays can be designed and used to study transcription factor activation in a variety of biological processes, including cell proliferation, differentiation, transformation, apoptosis, drug treatment, and others described herein.

Numerous methods have been developed for attaching hybridization probes to solid supports in order to perform immobilized hybridization assays and detect target oligonucleotides in a sample. Numerous methods and devices are also known in the art for detecting the hybridization of a target oligonucleotide to a hybridization probe immobilized in a region of the array. Examples of such methods and device for forming arrays and detecting hybridization include, but are not limited to those described in U.S. Patent Nos. 6,197,506, 6,045,996, 6,040,138, 5,424,186, 5,384,261, each of which are incorporated herein by reference.

Several modifications may be made to the hybridization arrays known in the art in order to customize the hybridization arrays for use in detecting activated transcription factors through the characterization of isolated transcription factor probes which form a complex with the activated transcription factors.

Since the hybridization probe arrays of the present invention are designed to hybridize to the probes in the transcription factor probe library by comprising a sequence that is complementary to the transcription factor recognition sequence, the composition of the hybridization probes in the array should complement the recognition sequences of the probes in the transcription factor probe library. As

discussed in Section 1, a variety of different libraries of transcription factor probes are provided that may be used to detect activated transcription factors according to the present invention.

5 Selection of the sequences used in the hybridization probes may be based on the different transcription factors that one wishes to detect in a sample. This, in turn, may depend on the type of organism, cell, or disease state one wishes to identify and/or monitor the gene expression of.

10 A significant feature of the present invention is the ability to detect multiple different transcription factors at the same time. This ability arises from the number of different DNA recognition sequences used in a library, the number of different DNA recognition sequences relating directly to the number of different transcription factors that the library can be used to detect. A given array of hybridization probes preferably has complements for at least 2, 3, 5, 10, 20, 30, 50, 100, 250 or more different DNA recognition sequences. The upper limit on the number of different  
15 DNA recognition sequences that the array of hybridization probes may detect is limited only by the number of known DNA recognition sequences and hence the number of known complements to the DNA recognition sequences.

A given array of hybridization probes may be used to detect gene expression in a single type of cell or organism or may be used to detect gene expression in  
20 multiple different types of cells or organisms. When the array is designed for use with a library designed to detect gene expression in multiple different types of cells or organisms, the array may include complements to DNA recognition sequences for multiple different types of cells or organisms. For example, the array may include complements to DNA recognition sequences for 2, 3, 4, 5 or more different types of  
25 cells or organisms. In one embodiment, the array may include complements to DNA recognition sequences for 10, 20, 30, 50, or more different types of cells or organisms.

If the sample to be analyzed comprises cells that may be from one or more different organisms, the DNA recognition sequences used in the library may be for  
30 all or some of the different transcription factors expressed by the one or more different organisms. For example, if library is to be used to classify an unknown

type of bacterium, the library may include DNA recognition sequences for multiple different types of bacteria, thereby allowing the library to be used to classify the bacterium. Accordingly, the array used in combination with the library would include complements to DNA recognition sequences for multiple different types of  
5 bacteria.

If the sample comprises cells from a particular organism, the DNA recognition sequences used in the library may be for all or some of the different transcription factors expressed by organism. Accordingly, the array used in combination with the library would include complements to DNA recognition  
10 sequences for the different transcription factors expressed by organism.

If the library is to be used to classify an unknown type of cells (i.e., determine whether a growth is malignant), the library may include DNA recognition sequences for multiple different types of cells including the different types of malignant, benign, and normal cell types present in the organism. Accordingly, the  
15 array used in combination with the library would include complements to DNA recognition sequences for the multiple different types of cells.

If the sample comprises cells of a single cell type, the DNA recognition sequences used in the library may be for all or some of the different transcription factors expressed by that cell type. Accordingly, the array used in combination with  
20 the library would include complements to the recognition sequences for all or some of the different transcription factors expressed by that cell type.

#### **i. Procedure for Performing Hybridization Using Array**

Provided below is a description of a procedure that may be used to hybridize  
25 isolated transcription factor probes to a hybridization array. It is noted that the below procedure may be varied and modified without departing from other aspects of the invention.

An array membrane having hybridization probes attached for the transcription factor probes is first placed into a hybridization bottle. The membrane  
30 is then wet by filling the bottle with deionized H<sub>2</sub>O. After wetting the membrane, the water is decanted. Membranes that may be used as array membranes include any



membrane to which a hybridization probe may be attached. Specific examples of membranes that may be used as array membranes include, but are not limited to NYTRAN membrane (Schleicher & Schuell), BIODYNE membrane (Pall), and NYLON membrane (Roche Molecular Biochemicals).

5           5 ml of prewarmed hybridization buffer is then added to each hybridization bottle containing an array membrane. The bottle is then placed in a hybridization oven at 42°C for 2 hr. An example of a hybridization buffer that may be used is EXPHYP by Clontech.

10           After incubating the hybridization bottle, a thermal cycler may be used to denature the hybridization probes by heating the probes at 90°C for 3 min, followed by immediately chilling the hybridization probes on ice.

          The isolated probe - transcription factors complexes are then added to the hybridization bottle. Hybridization is preferably performed at 42°C overnight.

15           After hybridization, the hybridization mixture is decanted from the hybridization bottle. The membrane is then washed repeatedly.

          In one embodiment, washing includes using 60 ml of a prewarmed first hybridization wash which preferably comprises 2X SSC/0.5% SDS. The membrane is incubated in the presence of the first hybridization wash at 42°C for 20 min with shaking. The first hybridization wash solution is then decanted and the membrane washed a second time. A second hybridization wash, preferably comprising 0.1X SSC/0.5% SDS is then used to wash the membrane further. The membrane is incubated in the presence of the second hybridization wash at 42°C for 20 min with shaking. The second hybridization wash solution is then decanted and the membrane washed a second time.

25

## **ii. Procedure for Detecting Array Hybridization**

          The following describes a procedure that may be used to detect isolated transcription factor probes isolated on the hybridization array. It is noted that each  
30    membrane should be separately hybridized, washed and detected in separate

containers in order to prevent cross contamination between samples. It is also noted that it is preferred that the membrane is not allowed to dry during detection.

According to the procedure, the membrane is carefully removed from the hybridization bottle and transferred to a new container containing 30 ml of 1X blocking buffer. The dimensions of each container is preferably about 4.5" x 3.5", equivalent in size to a 200 µL pipette-tip container. Table 2 provides an embodiment of a blocking buffer that may be used.

TABLE 2

**1X Blocking Buffer:**  
Blocking reagent: 1%  
0.1M Maleic acid  
0.15M NaCl  
Adjusted with NaOH to pH 7.5

It is noted that the array membrane may tend to curl adjacent its edges. It is desirable to keep the array membrane flush with the bottom of the container.

The array membrane is incubated at room temperature for 30 min with gentle shaking. 1 ml of blocking buffer is then transferred from each membrane container to a fresh 1.5 ml tube. 3 µl of Streptavidin-AP conjugate is then added to the 1.5 ml tube and is mixed well. The contents of the 1.5 ml tube is then returned to the container and the container is incubated at room temperature for 30 min.

The membrane is then washed three times at room temperature with 40 ml of 1X detection wash buffer, each 10 min. Table 3 provides an embodiment of a 1X detection wash buffer that may be used.

TABLE 3

**1 X Detection wash buffer:**  
10 mM Tris-HCl, pH 8.0  
150 mM NaCl  
0.05% Tween-20

30 ml of 1X detection equilibrate buffer is then added to each membrane and the combination is incubated at room temperature for 5 min. Table 4 provides an embodiment of a 1X detection equilibrate buffer that may be used.

**TABLE 4**

**1 X Detection equilibrate buffer:**

0.1 M Tris-HCl pH 9.5

0.1 M NaCl

The resulting membrane is then transferred onto a transparency film. 3 ml of CPD-*Star* substrate, produced by Applera, Applied Biosystems Division, is then pipetted onto the membrane.

A second transparency film is then placed over the first transparency. It is important to ensure that substrate is evenly distributed over the membrane with no air bubbles. The sandwich of transparency films are then incubated at room temperature for 5 min.

The CPD-*Star* substrate is then shaken off and the films are wiped. The membrane is then exposed to Hyperfilm ECL, available from Amersham-Pharmercia. Alternatively, a chemiluminescence imaging system may be used such as the ones produced by ALPHA INNOTECH. It may be desirable to try different exposures of varying lengths of time (e.g., 2–10 min).

The hybridization array may be used to obtain a quantitative analysis of the amount of transcription factor probe present. For example, if a chemiluminescence imaging system is being used, the instructions that come with that system's software should be followed. If Hyperfilm ECL is used, it may be necessary to scan the film to obtain numerical data for comparison.

**iii. Normalization of Data From Array Hybridization**

One of the advantages provided by array hybridization for detecting isolated transcription factor probes is the ability to simultaneously analyze whether multiple different activated transcription factors are present.

A further advantage provided is that the system allows one to compare a quantification of multiple different activated transcription factors between two or more samples. When two or more arrays from multiple samples are compared, it is desirable to normalize them.

5 In order to facilitate normalization of the arrays, an internal standard may be used so that the intensity of detectable marker signals between arrays can be normalized. In certain instances, the internal standard may also be used to control the time used to develop the detectable marker.

10 In one embodiment, the internal standard for normalization is biotinylated DNA which is spotted on a portion of the array, preferably adjacent one or more sides of the array. For example, biotin-labeled ubiquitin DNA may be positioned on the bottom line and last column of the array. In order to normalize two or more arrays for comparison of results, the exposure time for each array should be adjusted so that the signal intensity in the region of the biotinylated DNA is approximately  
15 equivalent on both arrays.

#### **7. Use of Multiple Libraries In Combination To Compare Gene Expression Between Different Samples**

20 When an array format is used for detecting the isolated probes, it may be desirable to use multiple libraries labeled with different detectable markers in order to facilitate comparison between samples. For example, as shown in Figure 3, probes from a first sample (e.g., a control sample) may have a green dye, and probes from a second sample (e.g., a test sample) may have a red dye. Both probes are  
25 separated from their bound complexes, mixed, and hybridized to a single array. Green spots in the array represent genes which only the cells in the control sample are expressing, and red spots in the array represent genes which only cells in the test sample are expressing. When both dyes hybridize to the same spot in an equal amount, the balanced mixture of green and red appears as yellow in the array,  
30 representing genes which cells in both the control and test samples are expressing.

One embodiment of this application of the present invention thus relates to a method for comparing gene expression between a test sample and a control sample,

the method comprising forming transcription factor probe - activated transcription factors complexes using the test sample and a first library of transcription factor probes having a first detectable marker; forming transcription factor probe - activated transcription factors complexes using the control sample and a second library of transcription factor probes having the same nucleic acid sequences as the first library but having a second, different detectable marker; isolating the transcription factor probes from the first library which formed complexes involving transcription factors from the test sample; isolating the transcription factor probes from the second library which formed complexes involving transcription factors from the control sample; and detecting the isolated transcription factor probes from the first and second libraries using a same hybridization array.

Another embodiment of this application of the present invention relates to a kit comprises the first and second library of transcription factor probes. The kit may optionally further include a hybridization array comprising complements to the transcription factor probes. The kit may also include instructions for isolating the transcription factor probe - activated transcription factor complexes using agarose gel.

## **8. Applications For Monitoring Gene Expression Via Detection Of Activated Transcription Factors**

By better understanding which cells express which genes and how different conditions influence gene expression, fundamental questions of biology can be answered. Thus, by being able to rapidly and efficiently detect multiple activated transcription factors at the same time, the present invention avails itself to numerous valuable applications relating to the monitoring of gene expression. Some of these applications are described herein. Other applications will be apparent to those of ordinary skill.

**a. Characterization of Cell Type**

By detecting and optionally quantifying which activated transcription factors are present in a cell sample, the methods of the present invention allow one to  
5 identify which genes are being expressed and to what extent each gene is being expressed. Different types of cells for a particular organism will express different genes. As a result, the present invention allows one to rapidly characterize a cell type based on which activated transcription factors are present and at what levels.

One embodiment of this application of the present invention thus relates to a  
10 method for characterizing a cell type, the method comprising forming transcription factor probe - activated transcription factors complexes using a test sample and a library of transcription factor probes comprising recognition sequences characteristic of different types of cells; isolating the transcription factor probes from the library which formed complexes involving transcription factors from the test sample; and  
15 detecting the isolated transcription factor probes using a hybridization array comprising sequences complementary to the transcription factor probes in the library.

A further embodiment of this application of the present invention relates to a library of transcription factor probes and hybridization array comprising  
20 compliments to the library of transcription factor probes are provided where the transcription factor probes comprise recognition sequences from multiple different cell types. A kit is also provided that comprises both the library of probes and the hybridization array. The kit may also include instructions for isolating the transcription factor probe – activated transcription factor complexes using an  
25 agarose gel, either in combination with the library, the hybridization array, or both.

It is noted that different organisms will also express different activated transcription factors. Characterizing the mixture of different activated transcription factors expressed by a particular organism (e.g., a culture of bacteria) can be used to identify the particular organism. This application of the method of the present  
30 invention may be particularly useful for rapidly characterizing microbes such as bacteria and diseased tissue.



One embodiment of this application of the present invention thus relates to a method for characterizing an organism, the method comprising forming transcription factor probe - activated transcription factors complexes using a test sample and a library of transcription factor probes comprising recognition sequences characteristic of different organisms; isolating the transcription factor probes from the library which formed complexes involving transcription factors from the test sample; and detecting the isolated transcription factor probes using a hybridization array comprising sequences complementary to the transcription factor probes in the library.

A further embodiment of this application of the present invention relates to a library of transcription factor probes and hybridization array comprising compliments to the library of transcription factor probes are provided where the transcription factor probes comprise recognition sequences from multiple different organisms. A kit is also provided that comprises both the library of probes and the hybridization array. The kit may also include instructions for isolating the transcription factor probe - activated transcription factor complexes using an agarose gel, either in combination with the library, the hybridization array, or both.

It is noted that the mixture of different activated transcription factors expressed by different cell types or organisms may be used according to the present invention as a form of an expression signature for that cell type. Figure 2 illustrates an array detection format. The pattern formed by the detectable markers [cubes] in Figure 2 can be used as a visual fingerprint of the expression signature and can be used to identify a particular cell type or organism based on that visual fingerprint. In this regard, it is envisioned that an array may be developed with a great multiplicity of immobilizing agents for different transcription factor probes. It is noted that the number of cubes shown in the figure is employed to reflect signal intensity.

The array may include immobilizing agents for transcription factor probes for different cell types and/or for different organisms. By comparing the array pattern to a standard for a particular cell type or organism, the cell type or organism can be rapidly determined.

**b. Determining the Functions of Different Genes**

Despite the fact that each cell in the human body contains the same set of genes, the human body is comprised of a wide diversity of different cell types that work in concert to form the human body. The wide diversity of cell types present in the human body and other multicellular organisms is due to variations between cells regarding which genes are expressed, the level at which the genes are expressed, and the conditions under which the genes are expressed. The present invention provides the unique ability of rapidly determining which of a great number of genes are expressed by numerous different cell types. By being able to determine which genes are expressed by which cell types, the functions of different genes can be deduced.

**c. Diagnosis of Disease States**

Certain disease states may be caused and/or characterizable by certain genes being expressed or not expressed as compared to normal cells. Other disease states may result from and/or be characterizable by certain genes being transcribed at different levels as compared to normal cells.

By being able to rapidly monitor the expression levels of multiple different genes, the present invention provides an accurate method for diagnosing certain disease states known to be associated with the expression non-expression, reduced expression, and/or elevated expression of one or more genes. Conversely, by comparing the expression non-expression, reduced expression, and/or elevated expression of one or more genes in normal and abnormal cells, present invention facilitates the association of one or more genes with certain disease states. By understanding that a particular disease state is caused by a different expression (higher or lower) of one or more proteins, it should be possible to remedy the disease state by increasing or decreasing the expression of the one or more proteins, by administering the one or more proteins or, if particular proteins are overexpressed, by inhibiting the one or more proteins.

One embodiment of this application of the present invention thus relates to a method for diagnosing a disease state of a sample of cells, the method comprising forming transcription factor probe - activated transcription factors complexes using the sample of cells and a library of transcription factor probes comprising different transcription factor recognition sequences; isolating the transcription factor probes from the library which form complexes involving transcription factors from the sample; detecting the isolated transcription factor probes using a hybridization array comprising sequences complementary to the transcription factor probes in the library; and diagnosing a presence of a disease state based on which transcription factors are activated in the cell sample as identified by which transcription factor probes are isolated.

#### **d. Compound Screening For Drug Candidates**

Being able to monitor transcription factor activity for multiple different transcription factors at the same time is of great importance to developing a better understanding of different roles that various transcription factors play. In addition, monitoring multiple different transcription factors at the same time allows one to rapidly screen for compounds that influence transcription factor activity, referred to herein as a "transcription factor modulator."

The present invention may thus be used as a high throughput screening assay for transcription factor modulators that either up- or down-regulate genes by influencing the synthesis and activation of transcription factors for those genes.

By having a further understanding of what compounds modulate transcription factor activity, such compounds may be more effectively used for *in vitro* modification of signal transduction, transcription, splicing, and the like, e.g., as tools for recombinant methods, cell culture modulators, etc. More importantly, such compounds can be used as lead compounds for drug development for a variety of conditions, including as antibacterial, antifungal, antiviral, antineoplastic, inflammation modulatory, or immune system modulatory agents. Accordingly, being able to monitor transcription factor activity for multiple different factors has

great use for screening compounds to identify lead compounds for pharmaceutical or other applications.

Indeed, because gene expression is fundamental in all biological processes, including cell division, growth, replication, differentiation, repair, infection of cells, etc., the ability to monitor transcription factor activity and identify compounds which modulate their activity can be used to identify drug leads for a variety of conditions, including neoplasia, inflammation, allergic hypersensitivity, metabolic disease, genetic disease, viral infection, bacterial infection, fungal infection, or the like. In addition, compounds which specifically target transcription factors in undesired organisms such as viruses, fungi, agricultural pests, or the like, can serve as fungicides, bactericides, herbicides, insecticides, and the like. Thus, the range of conditions that are related to transcription factor activity includes conditions in humans and other animals, and in plants, e.g., for agricultural applications.

As used herein, the term "transcription factor modulator" refers to any molecule or complex of more than one molecule that affects the regulatory region. The present invention contemplates screens for synthetic small molecule agents, chemical compounds, chemical complexes, and salts thereof as well as screens for natural products, such as plant extracts or materials obtained from fermentation broths. Other molecules that can be identified using the screens of the invention include proteins and peptide fragments, peptides, nucleic acids and oligonucleotides (particularly triple-helix-forming oligonucleotides), carbohydrates, phospholipids and other lipid derivatives, steroids and steroid derivatives, prostaglandins and related arachadonic acid derivatives, etc.

Existing methods for monitoring gene expression typically monitor downstream expression processes by measuring mRNA or the resulting gene product. However, why a particular mRNA or protein is expressed at higher or lower levels is not revealed by these methods. This is because a given compound can influence the formation of a transcription factor, influence the activation of the transcription factor, interact with the activated transcription factor, interact with the regulatory element to which the transcription factor binds, or interact with the mRNA that is produced.

By contrast, because the present invention is specific to detecting activated transcription factors, the present invention can be effectively used to screen for drugs that have a mechanism of action directly related to the expression and/or activation of transcription factors.

5 It should be noted that methods exist for measuring a transcription factor in a sample. However, because such methods detect the protein itself, they are unable to determine whether the transcription factor is activated, i.e., it is capable of binding to a regulatory element. By being able to detect whether multiple different transcription factors are activated, the present invention, when used in combination  
10 with an assay for detecting the amount of activated and unactivated transcription factor, allows one to evaluate specifically how a given compound influences the activation of different transcription factors.

The present invention may be used to screen large chemical libraries for modulator activity for multiple different transcription factors. For example, by  
15 exposing cells to different members of the chemical libraries, and performing the methods of the present invention, one is able to screen the different members of the library relative to multiple different transcription factors at the same time.

It will be appreciated that there are many suppliers of chemical compounds, including Sigma (St. Louis, Mo.), Aldrich (St. Louis, Mo.), Sigma-Aldrich (St.  
20 Louis, Mo.), Fluka Chemika-Biochemica Analytika (Buchs Switzerland) and the like.

In one preferred embodiment, high throughput screening involves testing a combinatorial library containing a large number of potential modulator compounds. A combinatorial chemical library may be a collection of diverse chemical  
25 compounds generated by either chemical synthesis or biological synthesis, by combining a number of chemical "building blocks" such as reagents. For example, a linear combinatorial chemical library such as a polypeptide library is formed by combining a set of chemical building blocks (amino acids) in every possible way for a given compound length (i.e., the number of amino acids in a polypeptide  
30 compound). Millions of chemical compounds can be synthesized through such combinatorial mixing of chemical building blocks.

Such combinatorial libraries are then screened to identify those library members (particular chemical species or subclasses) that modulate one or more transcription factors. The compounds thus identified can serve as conventional "lead compounds" or can themselves be used as potential or actual therapeutics for the one or more transcription factors whose activities the compounds modulate.

Preparation and screening of combinatorial libraries is well known to those of skill in the art. Such combinatorial libraries include, but are not limited to, peptide libraries (e.g., U.S. Pat. No. 5,010,175, Furka, *Int. J. Pept. Prot. Res.* 37:487-493 (1991) and Houghton et al., *Nature* 354:84-88 (1991)). Other chemistries for generating chemical diversity libraries can also be used. Such chemistries include, but are not limited to: peptoids (PCT Publication No. WO 91/19735), encoded peptides (PCT Publication WO 93/20242), random bio-oligomers (PCT Publication No. WO 92/00091), benzodiazepines (U.S. Pat. No. 5,288,514), diversomers such as hydantoins, benzodiazepines and dipeptides (Hobbs et al., *Proc. Nat. Acad. Sci. USA* 90:6909-6913 (1993)), vinylogous polypeptides (Hagihara et al., *J. Amer. Chem. Soc.* 114:6568 (1992)), nonpeptidal peptidomimetics with  $\beta$ -D-glucose scaffolding (Hirschmann et al., *J. Amer. Chem. Soc.* 114:9217-9218 (1992)), analogous organic syntheses of small compound libraries (Chen et al., *J. Amer. Chem. Soc.* 116:2661 (1994)), oligocarbamates (Cho et al., *Science* 261:1303 (1993)), and/or peptidyl phosphonates (Campbell et al., *J. Org. Chem.* 59:658 (1994)), nucleic acid libraries (see, Ausubel, Berger and Sambrook, all supra), peptide nucleic acid libraries (see, e.g., U.S. Pat. No. 5,539,083), antibody libraries (see, e.g., Vaughn et al., *Nature Biotechnology*, 14(3):309-314 (1996) and PCT/US96/10287), carbohydrate libraries (see, e.g., Liang et al., *Science*, 274:1520-1522 (1996) and U.S. Pat. No. 5,593,853), small organic molecule libraries (see, e.g., benzodiazepines, Baum *C&EN*, Jan 18, page 33 (1993); isoprenoids, U.S. Pat. No. 5,569,588; thiazolidinones and metathiazanones, U.S. Pat. No. 5,549,974; pyrrolidines, U.S. Pat. Nos. 5,525,735 and 5,519,134; morpholino compounds, U.S. Pat. No. 5,506,337; benzodiazepines, 5,288,514, and the like).

Devices for the preparation of combinatorial libraries are also commercially available (see, e.g., 357 MPS, 390 MPS, Advanced Chem Tech, Louisville Ky.,



Symphony, Rainin, Woburn, Mass., 433A Applied Biosystems, Foster City, Calif., 9050 Plus, Millipore, Bedford, Mass.). In addition, numerous combinatorial libraries are themselves commercially available (see, e.g., ComGenex, Princeton, N.J., Asinex, Moscow, Ru, Tripos, Inc., St. Louis, Mo., ChemStar, Ltd, Moscow, RU, 3D  
5 Pharmaceuticals, Exton, Pa., Martek Biosciences, Columbia, Md., etc.).

Control reactions may be performed in combination with the libraries. Such optional control reactions are appropriate and increase the reliability of the screening. Accordingly, in a preferred embodiment, the methods of the invention include such a control reaction. The control reaction may be a negative control  
10 reaction that measures the transcription factor activity independent of a transcription modulator. The control reaction may also be a positive control reaction that measures transcription factor activity in view of a known transcription modulator.

By being able to screen multiple different transcription factors at the same time, not only is it possible to screen a large number of potential transcription  
15 modulators per day, it is also possible to screen any potential transcription modulator relative to a large number of different transcription factors. The ability to screen multiple different transcription factors at the same time thus greatly enhances the high throughput capabilities of this screening assay.

#### 20           **d.       Evaluation of Drug Efficacy**

Given that certain disease states may be caused by an unusual level of transcription of one or more genes, drugs may be designed to either stimulate or inhibit transcription in order make gene expression of diseased cells approach the  
25 gene expression of normal cells. A rapid and effective method for monitoring gene expression is thus highly advantageous for evaluating the effectiveness of a drug's ability to alter the transcription of one or more genes. The effectiveness of a drug being delivered to a site of action as well as the drug's efficacy *in vivo* can thus be evaluated with the assistance of the methods of the present invention.

30           Also of great concern when developing new drugs is the side effects which the drugs might have. One approach for screening drug candidates for undesirable

side effects would be to employ the present invention to monitor how gene expression is altered in response to the administration of a drug candidate. By understanding how a candidate affects gene expression, candidates likely to have undesirable side effects can be rapidly identified.

5           Because the biological importance of transcription factors, they are ideal drug targets. Traditional transcription factor screening assays only detect one transcription factor at a time. As a result, existing assays are tremendously inefficient for detecting how a drug affects different gene expression. However, with the assistance of the present invention, it is now possible to screen hundreds and  
10 even thousands of transcription factors in a short amount of time in order to monitor how a given drug affects the expression of wide range of genes. The present invention will thus dramatically facilitate the screening process of identifying new drugs, characterizing their mechanism of the action, and screening for adverse side effects based on the drug's impact on expression.

15

**9.       Determining Sequence Binding Requirements For Transcription Factors**

          In general, a further application of the present invention is the rapid and  
20 efficient determination of the DNA sequence binding requirements for a given transcription factor. By being able to efficiently isolate DNA probes from multiple DNA probe – transcription factor complexes, the present invention makes it feasible to identify which DNA sequences bind to transcription factors, quantify the amount of each DNA sequence isolated, and use that information to determine different  
25 DNA sequence binding requirements for a given transcription factor in a high throughput manner.

          As a result of the completion of the human genome project, many new proteins will be discovered. Based on their primary structure, transcription factors for these proteins can be identified. However, the identification of DNA sequences  
30 to which these transcription factors will bind is a more difficult problem which the present invention helps to address.

When determining the DNA sequences to which transcription factors bind, there are several different questions that need to be answered. One question relates to the identification of an optimal binding sequence for the transcription factor. Another question relates to the identification of the minimal sequence required for binding. Yet another question which is related to the prior question, relates to the identification of a consensus sequence which is the minimum sequence required for binding. The present invention can be used to facilitate the answering of each of these questions.

The present invention enables one to rapidly determine a set of DNA sequences to which an activated transcription factor will bind and a set of DNA sequences to which an activated transcription factor will not bind. The amount isolated of each member in the set of DNA sequences to which an activated transcription factor binds may also be quantified.

Using this information, the minimum sequence required for the transcription factor to bind may be determined. Varying degrees of consensus sequences among the DNA sequences to which an activated transcription factor binds can also be determined. Meanwhile, by quantifying the amount of isolated DNA, how different base substitutions affect transcription factor binding can also be evaluated. This enables one to determine the sequences to which a given transcription factor binds most tightly.

Figure 4 illustrates a process whereby the minimum DNA sequence binding requirements for a given transcription factor can be rapidly determined. As illustrated, a sample containing an activated transcription factor is contacted with a library of probes. The probes in the library comprise a DNA sequence and a detectable marker.

Since, in this embodiment, the library of probes are used to determine the minimum DNA sequence binding requirements, the DNA sequences used in the library may be variations on a sequence which the transcription factor is known to bind to. Alternatively, the sequences used in the library may selected without knowledge of the binding specificity of the activated transcription factor.

If the DNA probes used to perform this method are known, a simple hybridization array having compliments to the DNA probes in the library may be employed, as described above. However, if a random library of DNA probes is employed, any isolated DNA probes can be characterized by existing position-fixed  
5 DNA array technology.

As a result of contacting the sample with the library of probes, complexes are formed between the activated transcription factor present in the sample and DNA probes in the library which have sequences that satisfy the sequence specificity of the DNA-binding domain of the activated transcription factor. The resulting probe –  
10 transcription factor complexes are preferably isolated by purification from agarose gel as described previously. After isolating the DNA probe – activated transcription factor complexes, those probes from the library which bind to transcription factors in the sample may be further isolated and characterized as discussed previously.

Since only those probes from the library which form a complex with an  
15 activated transcription factor will be isolated, identification of which probes are isolated serves to identify the range of sequences to which the activated transcription factor is capable of binding.

By constructing a consensus sequence based on the isolated probes, one is able to more precisely define the minimum binding requirements for the  
20 transcription factor. Furthermore, once a consensus sequence and a series of binding sequences are known, this information can be used to locate the occurrence of those sequences in 5' untranslation regions within a genome. This will allow researchers to identify which proteins may be regulated by the transcription factor. Genomes of different organisms may also be researched to identify proteins that may be  
25 functionally related.

Depending on the level of diversity of the library used, a further round of screening may be used to map the binding requirements of the transcription factor in greater detail. For example, when one or more binding sequences are identified, further experimentation may also be conducted to identify more binding sequences.  
30 This may involve creating a library of random mutations based on one or more DNA sequences shown to bind to the transcription factor in the prior screen. Binding of

the mutants may be performed in order to identify other mutants to which the transcription factor binds. Multiple cycles of generating and screening mutant libraries may be conducted as is necessary and desirable.

Figure 5 illustrates a variation of the method described in regard to Figure 4 where an optimal sequence for binding is identified. As illustrated in Figure 5, the isolated DNA probes are quantified as well as identified. By monitoring how changes in the sequence affect the amount of each probe isolated, one is able to see how the different sequences compete for binding to the transcription factor. As a result, an optimal transcription factor binding sequence can be identified.

Depending on the level of diversity of the library used to perform the first screen, when one or more binding sequences are identified, further experimentation may also be conducted to identify a better binding sequence. This may involve creating a library of random mutations based on one or more DNA sequences shown to bind to the transcription factor in the prior screen. Quantification of the binding of the mutants may be performed in order to identify a stronger binding mutant. Multiple cycles of generating and screening mutant libraries may be conducted as is necessary and desirable.

#### **10. Determining Transcription Factor Expression and Activation**

Prior to being activated, a transcription factor must be expressed. However, not all expressed transcription factors are activated. By determining whether multiple different transcription factors are being activated according to the present invention, in combination with determining whether the different transcription factors are being expressed, the present invention provides a rapid and efficient method for monitoring how different transcription factor expression and transcription factor activation change.

One application of this method relates to the diagnosis of disease states. By being able to track both changes in transcription factor expression and activation, more precise diagnosis of the cause of genetically related diseases may be discovered.

A further application of this method relates to the evaluation of how different agents or conditions affect both transcription factor expression and activation. For example, different agents can be screened for their ability to inhibit and/or activate the expression of certain proteins or impact different disease states. By knowing how the agent affects both transcription factor expression and activation, one is able to identify the mode of action of the agent. By being able to screen multiple transcription factors at the same time, one is further able to screen those agents for otherwise unforeseen adverse affects on the genetic level.

## 11. Kits For Use In The Present Invention

A wide variety of kits may be designed for use with the present invention. Various examples of these kits have already been described and additional kits are further described herein.

In one particular embodiment, a kit is provided which includes a library of transcription factor probes comprising recognition sequences for transcription factors. The kit further includes an array of hybridization probes where each probe comprises a sequence that is complementary to at least a portion of the recognition sequences on the probes in the library. In a preferred embodiment, the probes in the array of hybridization probes comprise 2, 3, 4, or more copies of a sequence that is complementary to at least a portion of the recognition sequences on the probes in the library.

In another particular embodiment, a kit is provided which includes a first library of transcription factor probes comprising recognition sequences for transcription factors, each probe in the first library further comprising a first detectable marker. The kit further includes a second library of transcription factor probes comprising the same recognition sequences for transcription factors as the first library, each probe in the second library further comprising a second detectable marker that is different than the first detectable marker.

In another particular embodiment, a kit is provided which includes a library of transcription factor probes comprising recognition sequences for transcription



factors in combination with instructions for using agarose gel to isolate transcription factor probe - transcription factor complexes.

5 In yet another embodiment, a kit is provided which includes an array of hybridization probes where each probe comprises a sequence that is complementary to at least a portion of recognition sequences of transcription factors. The kit further includes instructions for using agarose gel to isolate transcription factor probe - transcription factor complexes.

10 In yet another embodiment, a kit is provided which includes a hybridization array according to the present invention, hybridization buffer, detection wash buffer, and detection equilibrate buffer.

With regard to any of the kit embodiments, it is noted that the libraries of transcription factor probes and hybridization arrays may be varied as has been described herein.

## 15 12. Array Detection of Different Activated Transcription Factors

This section describes experimental results achieved by employing a library of probes according to the present invention to simultaneously screen samples of cells for 54 different activated transcription factors.

20 Figure 6 provides the sequences for the probes used to form the transcription factor probe library used in this experiment and the experiments described in Sections 13-19 herein. It is noted that the probes used were doubled stranded, Figure 6 only showing the strand with biotin labeled at the 5' end. The strands not shown are the complements to the sequences shown in Figure 6.

25 Figure 6 also shows the hybridization probes used in the hybridization probe array used in this experiment. Figure 7 shows the layout of the array of hybridization probes employed in the experiments described herein.

30 As can be seen in Figure 7, each hybridization probe was placed in multiple different regions, in this case, 4 separate regions. It is noted that 2, 3, 4 or more multiple separate regions may be employed. Alternatively, only a single region may also be employed. As can also be seen in Figure 7, the concentration of hybridization probe was varied in the different regions. This serves both as an

internal control as well as a mechanism for quantifying the amount of immobilized probes.

Biotinylated oligonucleotides which are used as controls were positioned in the regions of Row O and Column 17. These oligonucleotides also serve as a legend for the array, allowing a person to identify positions of rows and columns in the array.

The transcription factor probe library described in Figure 6 was contacted with the array of hybridization probes described in regard to Figures 6 and 7. In this instance, unlike the experiments to be described herein, no intermediate isolation of transcription factor probe - transcription factor complexes was performed.

Figure 8 is an image of the resulting array. As can be seen, all of the regions contain immobilized transcription factor probes. As can also be seen, the regions with higher concentrations of the same hybridization probe appear brighter (e.g., A1 and A2 vs. B1 and B2).

### **13. Array Detection of Selected Activated Transcription Factors**

Specific transcription factor probes and combinations of transcription factor probes were also contacted with a nuclear extract from HeLa cells and probes from any transcription factor probe - transcription factor complexes that formed being isolated. The transcription factor probes and combinations of transcription factor probes used were Brn3, c-Myb, Smad3/4 individually and the combination of Brn3, c-Myb, and Smad3/4.

Figures 9A-9D are images of the resulting arrays. As can be seen, only the hybridization probe regions for Brn3, c-Myb, and Smad3/4 appear to possess immobilized transcription factor probes in Figures 9A-9C respectively. Meanwhile, Figure 9D shows immobilized transcription factor probes in the Brn3, c-Myb, and Smad3/4 hybridization probe regions. The ratios of the spot densities among Brn3, c-Myb, and Smad3/4 has been found to be very similar to what is observed in arrays with single probe detection.

As can also be seen, the higher concentration regions appear consistently darker than the lower concentration regions. The use of regions with different

concentrations of hybridization probes allows different concentrations of transcription factor probes to be isolated. If the higher concentration region is saturated, the lower concentration one can be used for evaluation.

5     **14.     Array Detection of Activated Transcription Factors in HeLa Cells**

The entire library of probes described in regard to Figure 6 was also contacted with a nuclear extract from HeLa cells and probes from any transcription factor probe - transcription factor complexes that formed being isolated.

10         Figure 10B is an image of the resulting array. As a control, the entire library of probes described in regard to Figure 6 was also contacted with a control sample that did not contain any transcription factors. Probes from any transcription factor probe - transcription factor complexes that formed were isolated. Since no transcription factors are present in the control sample, it is expected that no  
15         complexes are formed and hence no probes are isolated. Figure 10A is an image of the resulting array.

As can be seen in Figure 10A and as would be expected, no transcription factor probes are immobilized in the array for the control sample. Meanwhile, as can be seen in Figure 10B, a myriad of transcription factor probes are immobilized  
20         in the array for the HeLa cell sample.

**15.     Comparison Between Activated Transcription Factors in HeLa Cells and PMA-activated HeLa Cells**

25         Figure 11A is an image of the array resulting from the entire library of probes described in regard to Figure 6 being contacted with a nuclear extract from HeLa cells and any probes from any transcription factor probe - transcription factor complexes that formed being isolated.

In comparison, Figure 11B is an image of the array resulting from the entire  
30         library of probes described in regard to Figure 6 being contacted with a nuclear extract from PMA-treated HeLa cells and probes from any transcription factor probe - transcription factor complexes that formed being isolated.

As can be seen by comparing Figures 11A and 11B, a number of transcription factors including Ets and NF-E1 can be seen to have been activated at higher levels in the PMA-treated HeLa cells.

5 The arrays shown in Figures 11A and 11B were imaged using a FluorChem imager (from Alpha Innotech Corp) in order to quantify the intensity of the spots appearing in the different regions of the array. Figures 12A and 12B provide tables of the signal intensity for the arrays shown in Figures 11A and 11B. Figure 12C provides the ratio between the intensities shown in Figures 12A and 12B.

10 As can be seen by comparing the data shown in Figures 12A and 12B, the signal intensities for the regions associated with transcription factors Egr (C5, C6, D5, D6); Ets (E9, E10, F9, F10); NF-E1 (G4, G5, H4, H5); and Smad3/4 (I15, I16, J15, J16) are more intense. Meanwhile, Ets (E9, E10, F9, F10) and Smad3/4 (I15, I16, J15, J16) show no difference in density.

15 The results of the experiment described in regard to Figures 11A, 11B, and 12A-12C was confirmed by performing a standard gel shift assay using Ets and NF-E1 probes. Specifically, nuclear extracts of HeLa and PMA-treated HeLa cells were incubated with Ets and NF-E1 probes respectively. As can be seen in Figure 13, bands corresponding to Ets appear for both PMA-treated and untreated HeLa cells. By contrast, a band corresponding to NF-E1 was not present in untreated HeLa cells  
20 but appeared in PMA-treated HeLa cells.

#### 16. Comparison Between Activated Transcription Factors in A431 Cells and PMA-activated A431 Cells

25 The entire library of probes described in regard to Figure 6 was also contacted with a nuclear extract untreated A431 cells and PMA-treated A431 cells. Figure 14A is an image of the array for untreated A431 cells and Figure 14B is an image of the array for PMA-treated A431 cells. As can be seen by comparing Figures 14A and 14B, transcription factors NF-E1 and NF-kB were found to be  
30 activated by PMA in A431 cells.

The results of the experiment described in regard to Figures 14A and 14B was confirmed by performing a standard gel shift assay using Ets, NF-E1, and NF-

kB probes. Specifically, nuclear extracts of A431 and PMA-treated A431 cells were incubated with Ets, NF-E1, and NF-kB probes respectively. As can be seen in Figure 15, bands corresponding to Ets appear for both PMA-treated and untreated A431 cells. By contrast, bands corresponding to NF-E1 and NF-kB were not present in untreated A431 cells but appeared in PMA-treated A431 cells.

#### 17. Comparison Between Activated Transcription Factors in Jurkat Cells and PMA-activated Jurkat Cells

The entire library of probes described in regard to Figure 6 was also contacted with a nuclear extract untreated Jurkat cells and PMA-treated Jurkat cells. Figure 16A is an image of the array for untreated Jurkat cells and Figure 16B is an image of the array for PMA-treated Jurkat cells. As can be seen by comparing Figures 16A and 16B, transcription factor AP1 was found to be activated by PMA. Interestingly, NF-E1 was not induced by PMA, showing that NF-E1 induction by PMA is cell line dependent.

#### 18. Comparison Of Activated Transcription Factors Between Multiple Cell Lines

The entire library of probes described in regard to Figure 6 was also contacted with nuclear extracts for HeLa, A431, Jurkat, K-562, and Y79 cells in order to compare the activated transcription factors present in these different cell lines. Figures 17A-17E show the resulting arrays for HeLa, A431, Jurkat, K-562, and Y79 cells respectively. As can be seen, the mixture of transcription factors that are activated varies for the different cells.

#### 19. Gel Shift Analyses Of Different Transcription Factors

Gel shift analyses were performed for multiple different transcription factors in order evaluate the sensitivity of gel shift analysis for detecting different transcription factors. Specifically, a nuclear extract of HeLa cells was incubated with transcription factor probes for c-Myb, Ets, Smad3/4, Brn3 and NF-E2. Figure

18A is an image of a polyacrylamide gel and Figure 18B is an image of an agarose gel. As can be seen in both figures, a band corresponding to probe – transcription factor complexes is effectively separated. As can also be seen, c-Myb, Ets, and Smad3/4 can be detected. However, Brn3 and NF-E2 are difficult to detect. By contrast, c-Myb, Ets, Smad3/4, Brn3 and NF-E2 can all be detected in the array shown in Figure 10B.

It will be apparent to those skilled in the art that various modifications and variations can be made in the compounds, compositions, kits, and methods of the present invention without departing from the spirit or scope of the invention. Thus, it is intended that the present invention cover the modifications and variations of this invention provided they come within the scope of the appended claims and their equivalents.